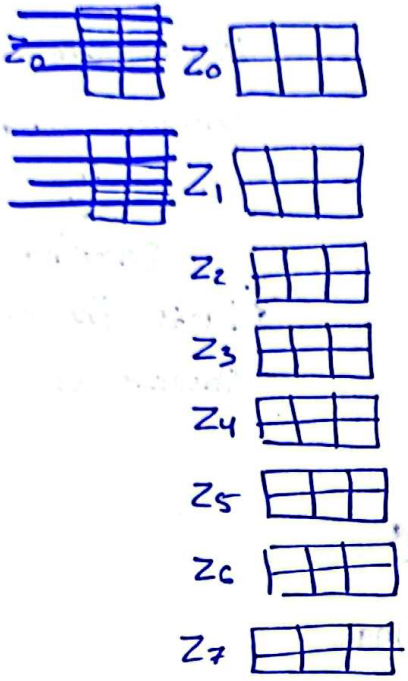
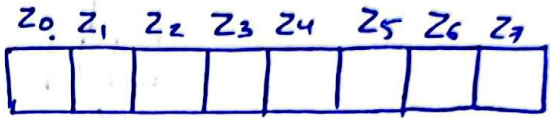
→ Feed Forward Neural Network

① Concatenate all attention heads





② Multiply with the weight matrix $\underline{\underline{W^0}}$ that was trained jointly with the model

③ The result would be $\underline{\underline{Z}}$ matrix that captures the information from all the attention heads. We can send this forward to FFNN.